Exercise 5.1: VB update equations (*19 points*)

In this exercise, you will derive variational Bayes (VB) update equations for the univariate Gaussian model defined by:

$$y = \mu + \varepsilon \tag{1}$$

$$p(\varepsilon) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau\varepsilon^2}{2}\right). \tag{2}$$

This model has two unknown parameters $\boldsymbol{\theta} = (\mu, \tau)$, where μ denotes the mean of the observations and τ the precision of the noise term.

In the following, expectation with respect to a probability distribution $q(x)$ is denoted by $E_x\{f(x)\} = \int f(x)q(x)dx$.

(a) (*2 points*) Assuming N independent observations $\boldsymbol{y} = (y_1, \ldots, y_N)^T$ are obtained, show that the log-likelihood for the model defined above is given by:

$$\log p(\boldsymbol{y}|\mu,\tau) = \frac{N}{2}\log\tau - \frac{\tau}{2}\sum_{n=1}^{N}(y_n-\mu)^2 - \frac{N}{2}\log 2\pi. \tag{3}$$

In this exercise, we will use a normal-gamma distribution – i.e. a combination of a Gaussian distribution over μ and a gamma distribution over τ – as prior over model parameters $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta}) = \sqrt{\frac{\lambda_0\tau}{2\pi}} \exp\left(-\frac{\lambda_0\tau}{2}(\mu-\mu_0)^2\right) \frac{b_0^{a_0}\tau^{a_0-1}}{\Gamma(a_0)} \exp(-b\tau) \tag{4}$$

Here, $\Gamma(x)$ denotes the so-called gamma function. Also note that the prior parameters λ_0, a_0, and b_0 must be positive.

(b) (*1 point*) Based on the log-likelihood function from Eq (3) and the prior from Eq (4), write down the log-joint distribution $\log p(\boldsymbol{y}, \boldsymbol{\theta})$ for the univariate Gaussian model.

Making use of the mean field approximation, we assume a factorization of the approximate posterior between mean μ and precision τ:

$$q(\boldsymbol{\theta}) = q(\mu)q(\tau) \tag{5}$$

(c) (*2 points*) Under the mean field approximation from Eq (5), show that the optimal variational distribution over μ is given by:

$$\log q^*(\mu) = -\frac{1}{2}E_\tau\{\tau\}\left(\sum_{n=1}^{N}(y_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right) + const \tag{6}$$

(d) (*2 points*) Show that the terms in Eq (6) can be reordered to give:

$$\log q^*(\mu) = -\frac{\bar{\tau}(N+\lambda_0)}{2}\mu^2 + \bar{\tau}\mu\left(\lambda_0\mu_0 + N\bar{y}\right) + const, \tag{7}$$

where $\bar{\tau} = E_\tau\left\{\tau\right\}$ and $\bar{y} = \frac{1}{N}\sum_{n=1}^{N} y_n$.

(e) (*2 points*) Compare Eq (7) to the logarithm of a Gaussian distribution. Show that $q^*(\mu)$ is a Gaussian with mean m and variance s^2 given by:

$$s^2 = \frac{1}{\bar{\tau}(N+\lambda_0)} \tag{8}$$

$$m = \frac{\lambda_0\mu_0 + N\bar{y}}{\lambda_0 + N} \tag{9}$$

(f) (*2 points*) Under the mean field approximation from Eq (5), show that the optimal variational distribution over τ is given by:

$$\begin{aligned}\log q^*(\tau) = &-\frac{\tau}{2}E_\mu\left\{\sum_{n=1}^{N}(y_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right\}\\ &-b_0\tau + \left(a_0 + \frac{N+1}{2} - 1\right)\log\tau + const\end{aligned} \tag{10}$$

(g) (*3 points*) Using the relation $E_\mu\left\{(x-\mu)^2\right\} = (x-m)^2 + s^2$, show that Eq (10) can be rewritten as:

$$\begin{aligned}\log q^*(\tau) = &-\frac{\tau}{2}\left(\sum_{n=1}^{N}(y_n-m)^2 + \lambda_0(\mu_0-m)^2 + (N+\lambda_0)s^2\right)\\ &-b_0\tau + \left(a_0 + \frac{N+1}{2} - 1\right)\log\tau + const\end{aligned} \tag{11}$$

(h) (*2 points*) Compare Eq (11) to the logarithm of a standard gamma distribution ($Gam(\tau|a,b) = \frac{b^a\tau^{a-1}}{\Gamma(a)}\exp\left(-b\tau\right)$). Show that $q^*(\tau)$ is given by a gamma distribution with parameters:

$$a = a_0 + \frac{N+1}{2} \tag{12}$$

$$b = b_0 + \frac{1}{2}\left(\sum_{n=1}^{N}(y_n-m)^2 + \lambda_0(\mu_0-m)^2 + (N+\lambda_0)s^2\right) \tag{13}$$

(i) (*3 points*) Show that the negative free energy can be expressed as:

$$\begin{aligned}\mathcal{F} = &-\frac{1}{2}E_\tau\left\{\tau\right\}\left(2b_0 + E_\mu\left\{\sum_{n=1}^{N}(y_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right\}\right)\\ &+\left(a_0 + \frac{N+1}{2} - 1\right)E_\tau\left\{\log\tau\right\} - \frac{N+1}{2}\log 2\pi - \log\Gamma(a_0)\\ &+ a_0\log b_0 + \frac{1}{2}\log\lambda_0 - E_\mu\left\{\log q(\mu)\right\} - E_\tau\left\{\log q(\tau)\right\}\end{aligned} \tag{14}$$

Exercise 5.2: VB implementation (*15 points + 1 bonus point*)

In this exercise, you will implement and test the variational Bayes update equations derived in the previous exercise. This exercise requires a computer.

(a) (*2 points*) Using the values $\mu = \tau = 1$, generate $N = 10$ observations from the univariate Gaussian model in Eq (1).

For the remainder of this exercise sheet, use the prior distribution from Eq (4) with parameters $\mu_0 = 0$, $\lambda_0 = 3$, $a_0 = 2$ and $b_0 = 2$.

(b) (*4 points*) Write a program that implements the update equations for μ (Eqs (8) and (9)) and τ (Eqs (12) and (13)). *Hint*: For a gamma-distributed random variable $\tau \sim Gam(\tau|a, b)$, the mean is given by $\bar{\tau} = E_\tau\{\tau\} = a/b$.

The expression for the free energy from Eq (14) can be simplified to:

$$\begin{aligned}\mathcal{F} = &-a \log b + \log \Gamma(a) - \log \Gamma(a_0) + a_0 \log b_0 + \frac{1}{2} \log \lambda_0 \\ &+ \log s - \frac{N}{2} \log 2\pi + \frac{1}{2} \end{aligned} \tag{15}$$

(c) (*2 points*) Extend your program to also evaluate the free energy $\mathcal{F}$ from Eq (15). *Hint*: For numerical stability, you should use a dedicated command to evaluate $\log \Gamma(\cdot)$ (e.g. `gammaln` in MATLAB), instead of evaluating $\Gamma(\cdot)$ first and then taking the log.

(c) (*2 points*) Extend your program to loop over the four update equations and the free energy. Monitor the free energy and stop once its difference between consecutive iterations falls below a predefined threshold (e.g.: 10^{-3}).

(d) (*5 points*) Run your program with the data generated in 5.2(a) and the prior parameters given above. To do this, you will also have to choose starting values for the posterior parameters m, s^2, a and b. Run the program once starting from the prior, i.e.: $m = \mu_0$, $s^2 = \frac{b_0}{a_0 \lambda_0}$, $a = a_0$ and $b = b_0$. Run the program a few more times starting from random values; although keep in mind that s^2, a and b must be positive. Compare the solution between the runs in terms of posterior parameter and free energy at convergence. If you can report the results from one run only, which one would you choose?

(e) (*1 bonus point*) Why would you choose $s^2 = \frac{b_0}{a_0 \lambda_0}$ for the initialization to the prior in 5.2(d) (and not $s^2 = 1/\lambda_0$)?

Bonus exercise 5.3: The true posterior (*5 bonus points*)

This is a bonus exercise. With the conjugate prior from Eq (4), it is possible to invert the univariate Gaussian model exactly. The true posterior is given by a normal-gamma distribution:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \overset{!}{=} \sqrt{\frac{\lambda\tau}{2\pi}} \exp\left(-\frac{\lambda\tau}{2}(\mu - m)^2\right) \frac{b^a \tau^{a-1}}{\Gamma(a)} \exp(-b\tau), \tag{16}$$

with parameters:

$$\lambda = \lambda_0 + N \tag{17}$$

$$m = \frac{\lambda_0 \mu_0 + N\bar{y}}{N + \lambda_0} \tag{18}$$

$$a = a_0 + \frac{N}{2} \tag{19}$$

$$b = b_0 + \frac{1}{2}\left(\frac{N\lambda_0}{N + \lambda_0}(\mu_0 - \bar{y})^2 + N\sigma_y^2\right) \tag{20}$$

where $\sigma_y^2 = \frac{1}{N}\sum_{n=1}^{N}(y_n - \bar{y})^2$.

(a) (*3 bonus points*) Write a program that creates a contour plot in the τ-μ-plane of the true posterior given in Eq (16) for the prior and data from 5.2(a).

(b) (*2 bonus points*) Extend the program to plot the approximate posterior from 5.2(d) as a contour plot into the same figure.

- Please submit your solutions via moodle `https://moodle-app2.let.ethz.ch/mod/assign/view.php?id=443089`.
- Post your questions to `https://moodle-app2.let.ethz.ch/mod/forum/view.php?id=441929`, or email *yao@biomed.ee.ethz.ch*.